Control Number :   GOOG-DOJ-AT-02512856

All Custodians :   Google

Custodian :   Google

Date/Time Created :   10/5/2021 12:47 AM

Date/Time Sent :

MODIFICATION DATE :   10/5/2021 12:47 AM

File Name :
                      Poirot_Modeling_Design_Doc.
                      docx

Poirot Modeling Design Doc
gTrade team

General Information

1. What does the model predict?

The model generates the bid reduction to be applied to DBM fixed CPM bids to maximize advertiser surplus.

2. How and where is the prediction used?

The bid reduction is applied on the fixed CPM bids before the internal DBM auction is run.

3. Contact information (team/TL email, wiki page).

gTrade. go/gtrade, gtrade-core@google.com

Source of Training Data

1. How is the training data obtained?

███

2. Does it correctly handle conversion data (see Conversion Handling)?

No conversion data used.

3. Does it correctly handle DBM data (see DBM Data Seperation)?

Yes. This pipeline only operates on DBM data.

4. How are bad data handled (e.g., when certain days of data are deemed unusable)?

We can easily exclude bad days through the borg files.

5. Are the training data stored?
    a. If so, does it satisfy wipeout compliance?

No user information is stored.

6. How frequently does the training data get refreshed?

Every day.

7. What happens when training data is stale?

We get alerts when any job files or the model file is stale.

Training Algorithm

1. How is the model trained?

Commented [1]: so this only works for fixed cpm? how about opt advertisers? and if not, why not?

Commented [2]: For opt, Ali wants to combine it with HDMI. We have to think about it. Fixed is 90% of DBM anyway, so we are ok with this in v1

Commented [3]: does this rely on running long standing holdback experiments?

Commented [4]: yes. we have to explore

Commented [5]: Eujin's suggestion: do this in the code directly rather than using an experiment framework
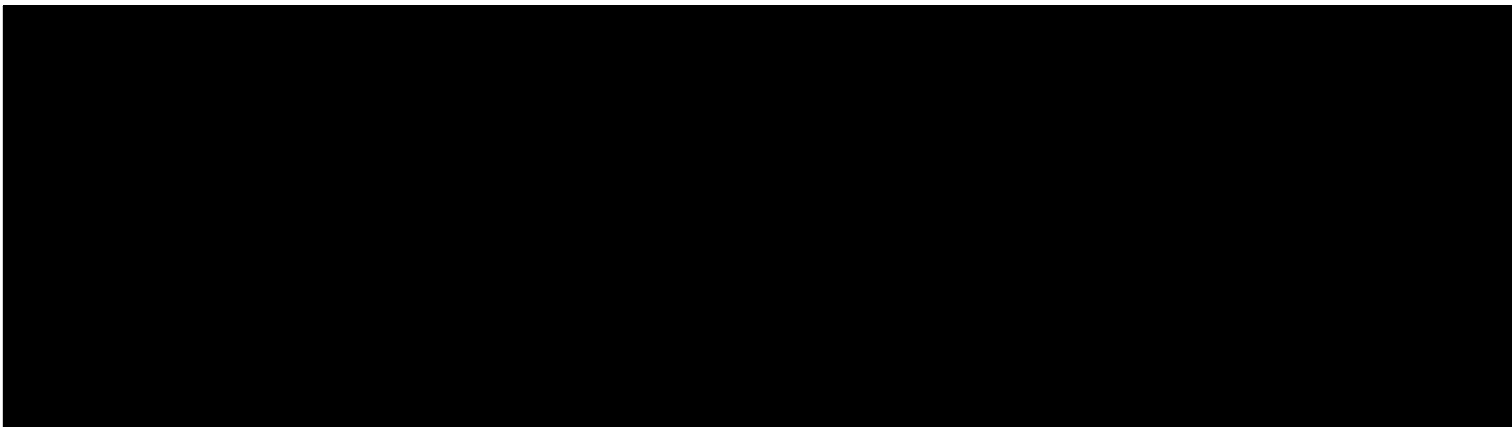
Commented [6]: why not use flogs? is there a migration plan?

Commented [7]: initially it was because we only had 20% data (that is no longer the case). We also have to flog a bunch of fields before we can switch to flogs.

Commented [8]: if you're relying on the filepusher alerts, those are somewhat unreliable.

Commented [9]: Do you have other suggestions? happy to incorporate

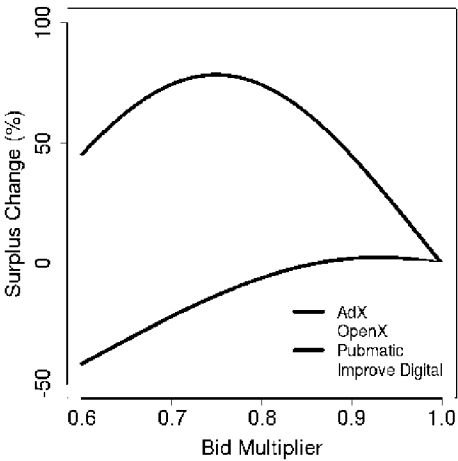**Commented [10]:** varz checker to check if the load time is recent enough

GOOG-DOJ-AT-02512856

Once the above model is fit separately for each exchange, we pick the *x* that maximizes surplus
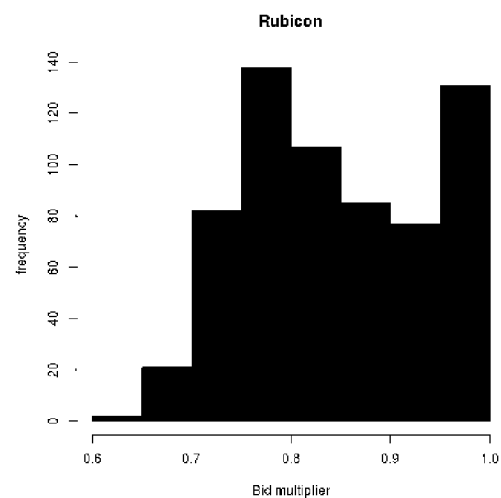for each advertiser.

The plot below shows surplus change vs bid multiplier (*x*) at the exchange level (i.e., ignoring
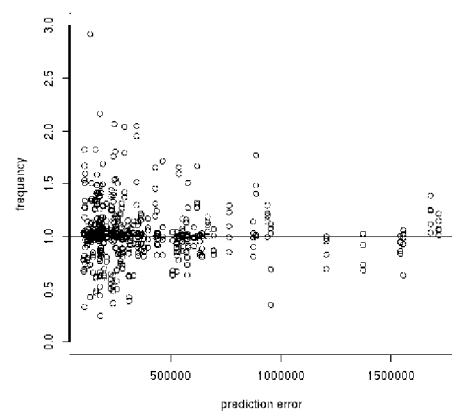the advertiser term) for 4 exchanges.

The plot below shows the histogram of bid multipliers at the advertiser level on the exchange
Rubicon.

Rubicon

Commented [12]: so one model per exchange?

Commented [13]: yes

Rubicon

Cross-validation

Mean squared error = 1.1, bias = 4%

Commented [14]: why is this a good accuracy metric?

Commented [15]: it sets a benchmark. it's not bad in accuracy + it seems to work well in experiments. For future launches, when we are adding features, we can check this metric to see if quality improves

Commented [16]: fix the label

   2. How frequently is the model updated?
Daily. Borgcron here.

   3. What is the metric for evaluating model quality?
Increase in surplus.

4. Are the models versioned?
Yes. Subscript vX to the model file names.

Serving

1. How is the model served?
Experiment file.

2. How does the model get updated in serving?
Dynamic files.

3. In case of production issues,
We can easily roll it back using an older experiment file.

4. Are there magic constants involved in serving (such as thresholds)?
If the surplus change is under 10%, we don't apply a bid multiplier

Monitoring

1. Which group is responsible for monitoring?
gtrade. gtrade-core@google.com

> **Commented [17]:** is there a pager on call rotation set up for this?
>
> **Commented [18]:** and how many people are responsible for fixing these problems?
>
> **Commented [19]:** it will be Tim for the most part unless he is not available, then i'll handle it. We don't have a pager for the analyst team (yet), but I heard Bahareh is working on it.
>
> **Commented [20]:** Add varzs in serving

2. (This is not an exhaustive list) Are there alerts for
    a. Training data looks bad and/or incomplete: Yes
        i. min # advertiser
        ii. min # exchanges
        iii. Min # background experiments
        iv. Min # impressions in the previous day's data
        v. Min value-added in the previous day's data
    b. Trained models have poor quality: Yes. cross-validation error is checked before model is pushed.
    c. Trained models are too different from the last iteration: Yes
    d. Models are stale - yes, email alerts
    e. Jobs failures (job down, borgcron overrun, borgcron failures) - yes, email alerts
    f. Server side monitoring (see productionisation guide for suggestion)
3. Is there a playbook describing how to handle the different types of alert?
    No

> **Commented [21]:** how bad are stale models?
>
> **Commented [22]:** I'm not terribly worried
>
> **Commented [23]:** why not?
>
> **Commented [24]:** Happy to add it

Production Readiness

1. How are PCRs handled?
We can easily switch to other clusters with placer deployment.

GOOG-DOJ-AT-02512859